

Université Claude Bernard LYON I

**43 Bd du 11 Novembre 1918
69622 Villeurbanne Cedex**

**I.N.R.A.
Institut National de la Recherche
Agronomique**

**Laboratoire d'Ecopathologie
63122 St Genes Champanelle**

D.E.A.

Analyse et modélisation des systèmes biologiques

Rapport Bibliographique

**Analyse exploratoire de tableaux de données
indicés dans le temps
en épidémiologie, en écologie et en agronomie**

**Emmanuel TILLARD
CIRAD-EMVT**

**Directeur scientifique: Bernard Faye
Parrains scientifiques: Daniel Chessel
Jean Thioulouse**

Février 1996

Sommaire

Introduction	1
I. Les analyses non structurées	4
I.1. Les analyses séparées par dates	4
I.2. Les analyses mélangées par dates	4
I.3. La projection d'éléments supplémentaires	5
I.4. Les classifications	6
II. Les analyses structurées	8
II.1. Les analyses interclasses et intraclases	8
II.1.1. ACP	9
II.1.2. AFC	10
II.1.3. Extensions possibles de la méthode	11
II.2. Les analyses inter-intraclases	12
II.2.1. Les analyses de tableaux juxtaposés	12
II.2.2. Les analyses sur variables instrumentales	13
II.2.3. L'analyse discriminante	14
III. Les analyses conjointes de K tableaux	15
III.1. La méthode STATIS	15
III.2. L'analyse triadique partielle	17
III.3. Les analyses 3 modes	18
IV. Les analyses de données temporelles	20
IV.1. L'ACP sur données brutes	21
IV.2. L'analyse sur graphe de voisinage	21
IV.3. L'ACP sur les paramètres d'un modèle	22
Conclusion	24
Remerciements	25
Bibliographie	25

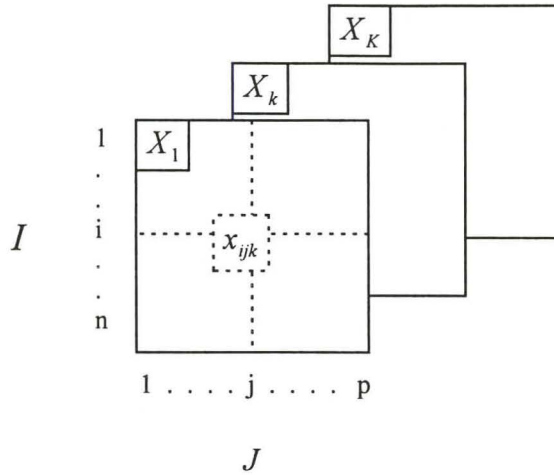
Introduction

Le champ d'investigation de l'écopathologie peut être cerné par les deux situations géographiquement opposées suivantes. Dans les pays développés, avec l'intensification de l'élevage, sont apparues de nouvelles manifestations pathologiques à caractère enzootique (qui s'expriment de façon continue ou répétitive dans les troupeaux), ne mettant plus en jeu exclusivement des contaminants spécifiques majeurs, mais relevant plutôt d'une inadaptation entre un ensemble de facteurs environnementaux (environnement géo-climatique, technico-économique), et les performances de production de l'animal (**Faye et al, 1991**). En France, ces pathologies sont les objets d'étude du Laboratoire d'Ecopathologie de l'I.N.R.A. (Institut National de Recherches Agronomiques). Dans les pays en voie de développement, l'élevage traditionnel est souvent confronté à des syndrômes pathologiques aux agents pathogènes mal caractérisés, très variables d'une exploitation à l'autre, d'une année à l'autre, suivant les conditions d'élevage ou climatiques, et pour lesquels des organismes de recherche-développement comme le CIRAD-EMVT (Centre International de Recherches Agronomiques pour le Développement - département Elevage et Médecine Vétérinaire des pays Tropicaux) essayent de trouver des plans de lutte adaptés. Aussi, les analyses de ces situations doivent-elles aborder le problème dans sa globalité, ou traduit en termes statistiques, sous ses aspects multidimensionnels, en prenant en compte des variables explicatives et expliquées multiples, structurées par le temps (**Faye et al, 1994b**).

En épidémiologie vétérinaire, la nécessité d'intégrer la dimension temporelle se traduit par la mise en place d'enquêtes longitudinales, basées sur des observations répétées sur les mêmes individus, qui permettent de prendre en compte plusieurs échelles de variation temporelle. Les données peuvent être indicées par un pas de temps réel déterminé, comme par exemple une saison, une année ou un groupe d'années (tendance séculaire) (**Faye et al, 1986b**). Elle peuvent être indicées également par l'âge des animaux, comme c'est le cas avec l'étude des courbes de croissance, où le poids des animaux est calculé pour différents âges types (**Moulin, 1993**). Enfin elles peuvent être indicées par cycle biologique et physiologique, comme le rang de lactation chez la vache laitière par exemple (**Faye et al, 1986a**), conduisant à la mise au point d'indicateurs longitudinaux caractérisant les carrières zootechniques et les biographies sanitaires (**Faye et al., 1994a**), éléments que l'on prendra de plus en plus en compte dans l'optimisation de la rentabilité des élevages (politique de réforme, choix du matériel génétique).

Le tableau de données typique généralement considéré à l'issue de ces enquêtes est un tableau ternaire X_{IJK} (un cube de données), croisant des objets I (des animaux ou des troupeaux) sur lesquels des variables J sont mesurées ou renseignées, à différentes dates, différents âges ou différentes périodes de temps K . On se ramène à un tableau binaire $I \times K$ (tableau d'évolution) si J n'est constituée que d'une seule variable (figure 1).

Figure 1: représentation des K tableaux X_{IJ} formant un cube de données de terme général x_{ijk} (avec $I(1 \rightarrow n)$, $J(1 \rightarrow p)$).



Le temps peut être considéré sous deux aspects qualitatifs différents. On peut l'envisager comme variable nominale, à modalités non séquencées à priori, par exemple dans l'étude de la variabilité inter-annuelle d'un phénomène, dans l'hypothèse d'indépendance des années entre-elles (**Faye et al, 1986b**). Le temps peut être également considéré comme une variable ordinale, pour laquelle les observations sont autocorrélées, comme par exemple dans l'étude des courbes de croissance (**Moulin, 1993**), les profils de notes d'état corporel (**Nicholson et Sayers, 1987**) et les profils de séroconversion (**Dessouter, 1995**).

L'étude des données indicées dans le temps peut être finalement abordé à différents niveaux de généralité décrit par **Pernin (1986)** selon l'objectif visé. Le niveau 1 correspond à la description d'une chronique univariée (un individu, une variable, observés au cours du temps), qui est généralement représentée à l'aide de courbes, comme par exemple l'évolution des incidences mensuelles des maladies d'élevage au cours de la lactation (**Faye et al, 1986a**). Ce niveau d'analyse relève des méthodes statistiques classiques, l'analyse des séries chronologiques, qui étudie en particulier l'autocorrélation entre observations successives, et l'analyse de variance qui compare des groupes d'individus entre eux. Les études à ce niveau de généralité semblent mal adaptées à la nature des objectifs préalablement cités, qui requièrent des analyses aux niveaux de généralité 2 (étude longitudinale uni-variable multi-individus et multi-variables uni-individu) ou 3 (étude des évolutions temporelles multi-variables multi-individus) (figure 2). De tels traitements appartiennent à la famille des

méthodes exploratoires multi-dimensionnelles, auxquelles nous limiterons ce travail. Cependant, le champ méthodologique ainsi circonscrit, il n'en existe pas moins une diversité de méthodes statistiques, reflétant à la fois la diversité des objectifs fixés et la diversité des disciplines les ayant développés (**Centofanti et al, 1989**). Aussi, dans le souci de prendre en compte un éventail assez large de méthodes, nous avons diversifier les sources de références à trois domaines principaux, l'épidémiologie, l'écologie et l'agronomie.

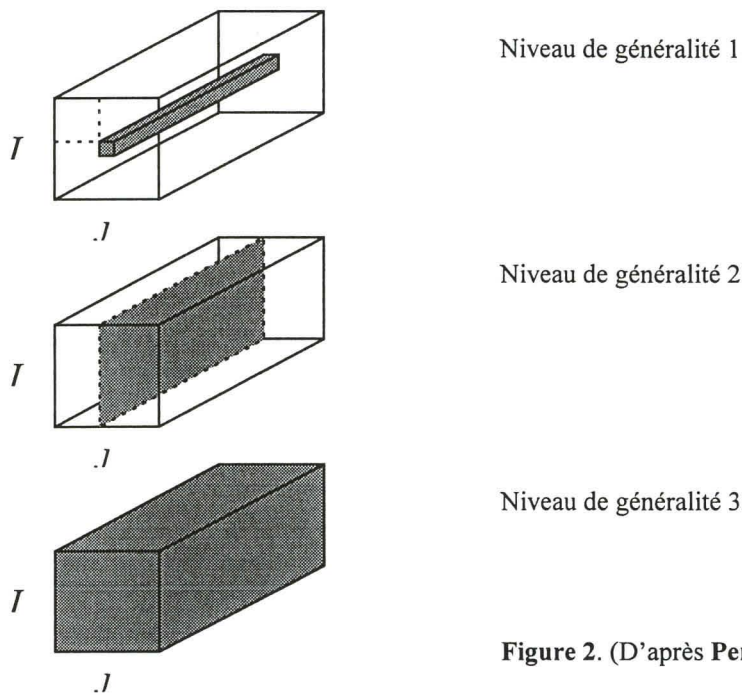


Figure 2. (D'après Pernin, 1986)

L'organisation de ce travail s'inspire de certaines revues méthodologiques (**Escoufier, 1980 et 1985 - Kiers, 1988 - Carlier et al, 1989 - Doledec et al, 1989**). Quatre parties seront successivement envisagées, selon une progression croissante dans la généralisation de la méthode. Nous décrirons dans un premier temps les analyses non structurées (séparées, mélangées), souvent associées à l'utilisation d'éléments supplémentaires et de classifications, puis les analyses structurées (inter et intraclasse, inter-intraclasse) dans lesquelles on associe au tableau initial des variables indicatrices représentant les différents effets que l'on désire prendre en compte ou éliminer, les analyses multi-tableaux, qui constituent une approche symétrique de l'ensemble des données, et les analyses de données temporelles dans lesquelles est prise en compte la notion de contiguïté entre observations. Un bilan de ces méthodes sera effectué en titre de conclusion.

I. Les analyses non structurées

Parmi les stratégies d'analyse décrites par **Williams et al. (1973)** et **Doledec et al. (1989)**, nous aborderons dans cette première partie les analyses factorielles séparées et mélangées, dans lesquelles la dimension temporelle n'est pas structurée. Les méthodes classiquement envisagées sont l'Analyse en Composantes Principales (ACP), l'Analyse Factorielle des Correspondances (AFC) et l'Analyse des Correspondances Multiples (ACM).

I.1. Les analyses séparées par date

Chacun des tableaux X_k est analysé séparément. Par exemple, **Faye et al. (1991)** analysent un tableau croisant J variables explicatives caractérisant la structure, les pratiques de conduite et les performances des animaux des exploitations de bovins laitiers I de 7 départements français, et ce pour 2 années consécutives, 1979 et 1980. Les centres de gravité des classes de variables obtenues pour les deux années après classification hiérarchique sont projetés sur le plan factoriel de l'analyse de l'année 1979. En comparant les classes par le nombre de modalités initiales communes, on met en évidence l'importance et la rapidité des évolutions qui marquent une population d'exploitations.

En écologie, **Furse et al. (1984)** comparent deux à deux des analyses séparées par des coefficients de corrélation établis entre les coordonnées factorielles des objets I (268 sites). 16 analyses différentes sont réalisées par saison (4) et par groupe taxonomique (4). Les 120 coefficients de corrélation obtenus pour chaque axe sont partitionnés en 3 groupes de comparaisons différents, les comparaisons inter-taxons (24), les comparaisons inter-saisons (24), et les autres comparaisons (72). Des tests t de Student permettent de comparer les moyennes des coefficients de corrélation des 3 groupes.

Quelque soit la méthode, graphique ou statistique utilisée pour comparer les tableaux X_k , l'analyse des évolutions est rendue difficile par l'absence de repère commun aux 2 analyses (**Faye et al, 1991**).

I.2. Les analyses mélangées par date

En écologie, le traitement de l'ensemble des observations s'impose pour comparer les deux types de variations, temporelle et stationnelle, et déceler une éventuelle inter-action

(Williams et al, 1973). Les K tableaux X_k sont transformés en un tableau Y ayant ik lignes et p colonnes. Gillet et al (1987), par exemple, effectuent une étude longitudinale des états de santé et des performances zootechniques dans des élevages de porcs en Bretagne. 29 élevages sont visités 4 fois sur une période de 8 ans. Lors de chaque visite, des mesures de performances sont effectuées et des questions sur la pathologie des animaux renseignées. Performances et états de santé ont été analysés séparément par AFC suivie d'une classification hiérarchique ascendante des lignes. Une variable à 4 modalités (une par date de visite) a été projetée en élément supplémentaire sur les cartes factorielles dans les 2 analyses pour visualiser l'évolution temporelle. Cette évolution est particulièrement nette pour l'analyse des états de santé, dans laquelle on constate une détérioration de l'état sanitaire des truies au fil des 4 passages. Cependant, les 2 typologies, mises en relation dans un tableau de contingence croisant les classes de santé avec les classes de performance, montre une association paradoxale entre dégradation de l'état sanitaire et évolution favorable des performances, qui pourrait illustrer ici le principal inconvénient de cette stratégie, qui réside dans le fait que les facteurs couplés (IK) ne sont pas dissociables, et que chacun d'entre eux peut masquer la structure de l'autre (Doledec et al, 1989).

En Hydrobiologie, Dakki (1986) illustre particulièrement bien cet inconvénient pour des données décrivant la variabilité spatio-temporelle de l'abondance des espèces de poissons d'un cours d'eau au Maroc. 6 relevés d'abondance sont effectués au cours d'une année, dans 14 stations réparties le long du cours d'eau. L'auteur utilise un tableau Y à 84 (14 stations x 6 dates) lignes et 48 colonnes (espèces) et montre que l'amplitude d'habitat des espèces est lié à un gradient thermique qui contient une part liée au gradient thermique altitudinal (effet station) et une part liée au gradient thermique saisonnier, que l'analyse ne permet pas de séparer.

L'analyse mélangées peut parfois surmonter cette confusion dans le cas de situation simple, par exemple lorsque la variation spatiale ne concerne que 2 sites (Swaine et al, 1980). Face à des situations plus hétérogènes, on est souvent conduit à essayer de séparer les deux effets, à l'aide de projection d'éléments supplémentaires ou de méthodes de classification.

I.3. La projection d'éléments supplémentaires

L'évolution temporelle est souvent visualisée par la projection d'éléments supplémentaires sur les cartes factorielles et par les trajectoires des lignes entre les différentes

dates d'observation. **Tourrand (1994)** réalise par exemple deux typologies d'exploitations d'élevage traditionnel dans la région du Delta du fleuve Sénégal, à 5 années d'intervalle, sur la base des mêmes variables J . Les 95 exploitations de 1985 se sont scindées en 128 exploitations en 1990 (par division de certaines des exploitations). Ces 128 exploitations sont projetées en éléments supplémentaires sur le plan factoriel 1×2 issu de l'analyse de 1985, et les trajectoires (mono ou bifides) sont visualisées. Les modalités de variables indicées dans le temps peuvent être également projetées en éléments supplémentaires comme le réalise **Lopez et al. (1991)** en projetant sur le plan factoriel 1×2 de l'AFC d'un tableau croisant 561 exploitations (I) et 33 modalités de variables caractérisant les pratiques d'élevage (J), 6 variables décrivant les taux cellulaires dans le lait à 6 dates différentes.

Cazes (1982, I et II) envisage, dans le cadre de tableaux ternaires X_{IJK} , 3 cas d'analyse: celle d'un tableau $X_{(IK)J}$ sur lequel on projette X_{KJ} (pour calculer des variances inter-temps), X_{IJ} (pour calculer des variances inter-objet), $X_{(IK)(JK)}$ pour représenter chaque j de J à chaque k de K , symétriquement celle d'un tableau $X_{I(JK)}$, et enfin l'analyse d'un tableau X_{IJ} sur lequel on projette $X_{(IK)J}$, $X_{I(JK)}$, X_{KJ} , et X_{IK} . Cette dernière analyse est une véritable analyse discriminante inter-catégorie (**Persat et al, 1989**). Deux exemples d'utilisation « massive » des éléments supplémentaires sont fournis par l'économétrie, dans le cas d'un tableau ternaire (**Gopalan, 1980**) et même quaternaire (**Mourad, 1983**).

I.4. Les classifications

Les classifications sont des méthodes qui prolongent classiquement la plupart des analyses factorielles vues précédemment. Nous n'envisagerons ici que quelques utilisations particulières, visant à séparer 2 effets croisés i et k sur la marge ligne d'un tableau, et mises en oeuvre indépendamment des méthodes factorielles.

En écologie, **Vincent (1983)** et **Vincent et al. (1983)** réalisent l'AFC d'un tableau d'abondance Y croisant 24 relevés (6 stations \times 4 dates) et 31 espèces macrobenthiques, et une classification des lignes dont ils matérialisent les groupes dans un tableau croisant les 4 saisons et les 6 stations. Ils démontrent clairement, en s'appuyant sur des représentations graphiques (dendrogramme, distribution spatio-temporelle) l'existence d'une importante discontinuité interclasse d'origine spatiale, et une variabilité saisonnière différente selon les stations (discontinuité intraclasse), confirmée par des analyses de variance.

En épidémiologie humaine, **Facy et al. (1984)** appréhendent la dynamique du développement de l'enfant de moins de 3 ans par des méthodes d'intersection de classifications. Ils analysent un tableau croisant 415 nourrissons, 19 variables décrivant la santé physique et psycho-affective, à 3 âges différents, 9, 18 et 36 mois. Pour chacun des 3 âges, une classification croisée est effectuée à la fois sur les lignes et les colonnes, en utilisant l'algorithme des nuées dynamiques. Les P classes sont ré-ordonnées suivant l'ordre croissant d'un critère de risque épidémiologique défini par $C_P = \sum_{i \in P} \sum_{m=1}^M \delta_i^m$, avec M le nombre de modalités de la classe P , et $\delta_i^m = 1$ si i a la modalité m , 0 sinon. On aboutit pour chaque âge à une partition des groupes de sujets à risque. Un tableau de contingence croise deux à deux les partitions obtenues pour chaque âge, et visualise l'évolution des enfants d'un groupe à risque à un autre (trajectoires). En épidémiologie vétérinaire, pour établir des trajectoires sanitaires à l'échelle de la carrière des vaches laitières, considérée comme une succession de lactations dans le temps, **Faye et al. (1994a)** réalise une classification hiérarchique ascendante des lignes d'un tableau croisant 3851 lactations, tous rangs confondus et 8 variables sanitaires, qui aboutit à l'individualisation de 6 groupes de lactations, caractérisés par un profil sanitaire. Pour une longueur de carrière déterminée (1 à 7 lactations), on construit un tableau croisant les vaches en lignes et les rangs de lactation en colonnes, de terme général le numéro de groupe (issu de la classification). Une seconde classification sur les lignes aboutit à une typologie des trajectoires sanitaires individuelles.

En écologie, **Kosmelj et al. (1983)** adaptent les méthodes de la classification aux séries chronologiques. Ils classent 9 séries chronologiques (attachées à 9 nappes phréatiques différentes) constituées de 52 observations hebdomadaires de 3 variables décrivant le débit du Rhône, la température et la pluviométrie, dont ils étudient l'impact sur les nappes. 2 distances originales entre séries sont proposées, selon la similarité de leurs corrélogrammes et leur corrélation croisée (fonctions spécifiques des séries).

Les analyses séparées et mélangées que nous venons de voir, enrichies de la pratique des éléments supplémentaires et des classifications, sont simples à mettre en oeuvre, et constituent une première approche. Cependant, ces analyses ne permettent pas de séparer de manière satisfaisante et d'analyser spécifiquement la composante temporelle, car le temps n'est pas directement pris en compte, mais simplement réintroduit au moment de l'interprétation. Les analyses structurées regroupent un ensemble de méthodes qui introduisent

dans les données la structure du plan d'expérience, et dont le principe consiste à rechercher la part de variabilité dépendant d'un facteur, choisi à priori, ainsi que la part qui en est indépendante, ou qui ne peut s'expliquer que par une interaction avec d'autres facteurs (Thioulouse et al, 1987).

II. Les analyses structurées

Dans ces analyses, on transforme le tableau ternaire en un tableau $X_{(IK)J}$ aux marges duquel on associe une ou plusieurs variables qualitatives complémentaires qui structurent le plan expérimental, notamment au niveau temporel. Ces variables peuvent engendrer plusieurs sous-espaces vectoriels orthogonaux entre eux, sur lesquels on projette $X_{(IK)J}$ (Doledec et al, 1987). La figure 3 schématise pour une variable du tableau initial X_1 , la décomposition de la variance sur des sous-espaces vectoriels engendrés par 2 variables complémentaires A_1 et A_2 . Ce chapitre se divisera en deux parties, les analyses factorielles (ACP et AFC) inter-classes et intra-classes, et les analyses inter-intra classes (prise en compte simultanément de deux effets différents).

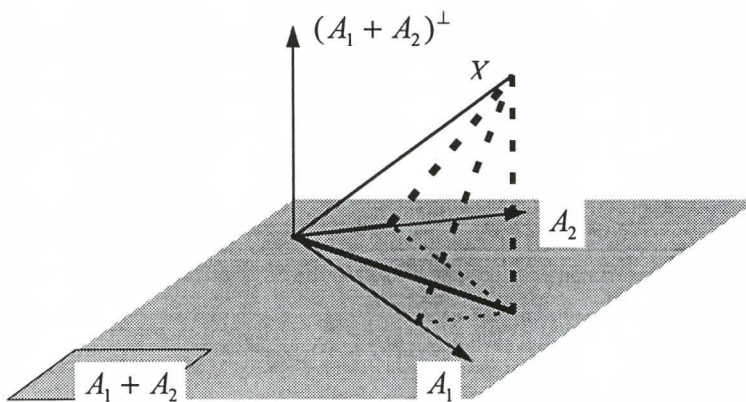


Figure 3: Schématisation de la décomposition de la variance de X_1 (norme au carré) dans les sous-espaces engendrés par les deux variables A_1 et A_2 (traits pleins), et dans les sous-espaces qui leur sont orthogonaux (traits en pointillé). On retrouve ainsi les 6 situations de Doledec et al, 1987.

II.1. Les analyses inter-classes et intra-classes

Ces analyses (ACP ou AFC) sont particulièrement adaptées à la prise en compte ou à l'élimination d'effets simple, comme l'effet saisonnier. Elles reposent sur une décomposition de la variance totale du tableau en une variance intraclasse et une variance interclasse, matérialisées par la figure 3 (Doledec et al, 1987 - Escofier et al, 1983 - Doledec et al, 1989).

II.1.1. ACP

Dolédéc et al. (1987) envisagent l'analyse en composantes principales d'un tableau $X_{(IK)J}$ croisant 24 lignes (6 stations x 4 dates) et 10 variables quantitatives décrivant la physico-chimie d'un cours d'eau, le Méaudret, sous divers points de vue. L'analyse classique (analyse mélangée du tableau $X_{(IK)J}$), décrit simultanément rythme saisonnier et typologie stationnelle. 6 autres points de vue sont considérés: pour chaque variable j envisagée, les données centrées forment un tableau d'analyse de variance à deux facteurs (stations et dates) sans répétition. On substitue au tableau initial des données centrées et réduites, des tableaux modèles, successivement:

X_1 , le tableau des moyennes par stations, et $X - X_1$ le tableau des écarts à ces moyennes
 X_2 , le tableau des moyennes par dates et $X - X_2$ le tableau des écarts à ces moyennes
 X_{12} , le tableau du modèle linéaire additif et $X - X_{12}$ le tableau des écarts à ces valeurs.

L'ACP interclasses est l'ACP des centres de gravité des classes (moyenne), alors que l'ACP intraclasses est l'ACP des écarts au centre de gravité des classes. D'un point de vue géométrique, chaque tableau X_1, X_2, X_{12} est le projeté de X sur un sous espace vectoriel défini respectivement par le vecteur « station », le vecteur « dates » et le sous espace vectoriel (stations + dates). De même, chaque tableau $X - X_1, X - X_2, X - X_{12}$, est le projeté de X sur un sous espace vectoriel orthogonal aux précédents vecteurs respectivement. L'ACP généralisée (sans centrage) des tableaux projetés permet de réaliser la typologie des relevés en prenant en compte de chaque point de vue:

X_1	:on analyse les différences entre stations (typologie stationnelle simplifiée)
$X - X_1$:on élimine les différences entre stations (typologie temporelle)
X_2	:on analyse les différences entre dates (typologie temporelle simplifiée)
$X - X_2$:on élimine les différences entre dates (typologie stationnelle)
X_{12}	:on tient compte de l'interaction stations - dates (structures espace-temps)
$X - X_{12}$:on élimine l'interaction stations - dates.

Les aspects temporels sont mieux analysés par $X - X_1$ que par X_2 car on conserve une image de la variabilité interne à la catégorie étudiée (information commune aux stations), qui renforce la définition des positions relatives des dates: la projection sur l'orthogonal du sous - espace station l'emporte sur la projection sur le sous - espace temps (**Dolédéc et al., 1987**). On note également que la procédure choisie contient deux projections successives, celle du point de vue envisagé (parmi les 6 présentés antérieurement) et celle de l'ACP. On

pourra donc étudier la décomposition de la variance entre analyses (point de vue) différentes, puis la décomposition de la variance propre à chaque analyse, sur chacun des axes. Avec des points de vue tout à fait similaires, les analyses inter-classes et intra-classes peuvent être étendues à l'analyse factorielle des correspondances.

II.1.2. AFC

Escofier et Drouet (1983) propose une méthode de comparaison de plusieurs tableaux de fréquence définis sur le même couple d'ensemble I et J , et une application comparant l'évolution sur onze années des causes de mortalité par classes d'âge (**Drouet, 1983**). $X_{(ik,j)}$ est un tableau à ixk lignes et j colonnes, de terme général x_{ikj} , de somme marginale ligne $x_{ik.}$, de somme marginale colonne $x_{.j}$ et de somme totale $x_{...}$. Les lignes sont regroupées en k blocs (k sous-tableaux indicés dans le temps). $X_{(k,j)}^+$ est un tableau à k lignes et j colonnes de terme général $x_{kj}^+ = \sum_{i \in k} x_{ikj}$ et de somme marginale ligne $x_{k.}^+ = \sum_{j=1}^c x_{kj}^+$. La somme marginale colonne et la somme totale restent inchangées. On définit également un tableau $X_{(ik,j)}^-$ à ixk lignes et j colonnes, de terme général $x_{ikj}^- = x_{ikj} - \frac{x_{kj}^+ x_{ik.}}{x_{k.}^+} + \frac{x_{ik.} x_{.j}}{x_{...}}$. Ces modifications sont les versions AFC des centrages décrits plus haut en ACP. Les sommes marginales de $X_{(ik,j)}^-$ sont celles de $X_{(ik,j)}$. L'AFC inter-classes est définie par l'AFC de $X_{(k,j)}^+$, accompagnée de la projection en éléments supplémentaires des ixk lignes de $X_{(ik,j)}$, pour représenter l'écart des profils des ixk lignes à leur moyenne. Elle maximise la variance des centres de gravité des blocs. L'AFC intra-classes est l'AFC de $X_{(ik,j)}^-$. Elle maximise la moyenne de la variance interne aux blocs. Dans l'exemple de Drouet, les profils de mortalité j des différentes classes d'âge i d'une même année k diffèrent beaucoup plus entre eux que ne diffère le profil d'une même classe d'âge au cours du temps. Dans ce cas précis, seule l'analyse intra-classes traduira les différences entre lignes homologues (**Escofier et al, 1983 - Escofier et al, 1988b**). En écologie, ces méthodes d'ordination sous contraintes sont particulièrement intéressantes lorsqu'une structure forte, considérée comme triviale, les variations saisonnières par exemple, masque une structure plus faible, comme la structure longitudinale d'une rivière.

Cependant, les limites de la méthode sont précisées. Lorsque l'on ne centre plus l'analyse sur une typologie des colonnes (nombre de colonnes faible), on privilégiera la modélisation des tableaux espace - temps pour chacune des espèces, par un modèle soit additif, soit multiplicatif ou mixte, comme par exemple l'ACP-3 modes décrite par **Kroonenberg (1983)**, qui prend en compte les inter-action d'ordre 3 (espèces x stations x dates). Lorsque le nombre de dates (respectivement des stations) est faible, on s'orientera vers des analyses séparées par date (respectivement par station), en utilisant dans le second cas des méthodes liées à l'analyse des séries chronologiques (moyennes mobiles pour l'analyse des tendances et corrélogrammes pour l'analyses des variations cycliques).

II.1.3 . Extensions possibles de la méthode

La méthode décrite précédemment peut être généralisée à l'introduction d'une contrainte sur les deux marges d'un tableau (**Cazes et al, 1988**). Les auteurs analysent un tableau croisant 35 relevés (i) effectués dans 6 stations à 6 dates différentes (une valeur manquante) et 43 espèces (j) réparties en 4 groupes taxonomiques, l'objectif étant d'effectuer une typologie stationnelle en éliminant l'effet temporel (sur les lignes) et l'effet groupe taxonomique (sur les colonnes), pour mesurer la capacité des groupes à servir d'indicateurs biologiques, en relation avec les coûts élevés d'identification des espèces. L'analyse des correspondances internes (analyse intra-classes sur les deux marges) analyse un tableau $R_{(i,j)}$ à i lignes et j colonnes, de terme général $r_{ij} = P_{(i,j)}^\circ + p_i.p_{.j}$ avec $P^\circ = Q_X' P Q_Y$, double projection du tableau initial P sur le sous espace défini par les indicatrices de dates Q_X et les indicatrices de groupes taxonomiques Q_Y . La variance supportée par chacun des axes de l'ACI est décomposée par groupe des partitions imposées au tableau initial.

Franquet et al. (1994) élargit les analyses inter et intra-classes à l'analyse de co-inertie (analyse deux tableaux couplés), d'un tableau faunistique et d'un tableau mésologique, structurés par dates. Les mesures d'abondance de 5 taxons et de 3 variables de milieu ont été effectuée dans 12 relevés (3 dates x 4 stations.) L'objectif de l'étude est double: trouver des combinaisons de variables de chacun des tableaux exprimant une covariation temporelle (analyses inter-dates) et une costructure spatiale (analyses intra-dates), entre faune et milieu.

Les analyses que nous avons décrites jusqu'à présent apportent un réel progrès dans la prise en compte d'effets structurant les données. Cependant, elles ne permettent pas d'intégrer

plusieurs effets à la fois, comme le montrent les 6 points de vue différents envisagés par **Doledec et al. (1987)**. Des méthodes plus élaborées permettent d'associer dans une même analyse un ou plusieurs effets inter-classes, un ou plusieurs effets intra-classes, et leur interactions (**Doledec et al, 1989**). Nous envisagerons successivement l'analyse de tableaux juxtaposés, les analyses sur variables instrumentales, et la méthode LONGI.

II.2. Les analyses inter-intra classes

II.2.1. L'analyse de tableaux juxtaposés

En parallèle à l'analyse intra-classes et l'analyse inter-classes, **Escofier et al. (1988b)** proposent une méthode pour analyser des tableaux de fréquences juxtaposés, dérivée d'une méthode initialement conçue pour analyser des tableaux de dimensions trop grande (nombre trop important de colonnes). Les auteurs montrent que l'AFC d'un tableau $X_{(I,J,K)}$ sur lequel une contrainte est placée en colonne équivaut à l'ACP non réduite d'un tableau constitué des facteurs issus des AFC des sous tableaux $X_{(I,J)}$, préalablement « redressés » en multipliant leurs composantes par $f_{i..}/f_{i..}$, et des facteurs du tableau des marges (de terme général $x_{i..}$). Cette analyse tient compte ainsi de la dispersion inter-tableaux et intra-tableaux.

Deux exemples d'utilisation de cette méthode en épidémiologie sont fournis par **Faye et al. (1986a)**. On étudie 4 tableaux de fréquence juxtaposés dans le temps (4 années - 1978 à 1981), croisant en lignes 7 variables (en présence - absence), décrivant la pathologie individuelle des vaches laitières, en colonnes 20 variables décrivant l'exploitation dans laquelle se trouve l'animal (sous tableaux de Burt). Le plan factoriel 1x3 individualise les années 1979-1981 de l'année 1978. L'examen des cartes factorielles annuelles permet de mieux expliquer les variations inter-annuelles. Dans un autre article **Faye et al. (à paraître)**, étudient les trajectoires sanitaires à l'échelle de la carrière des vaches laitières (succession dans le temps de différentes lactations). Pour une succession de k lactations déterminées (k pouvant varier de 1 à 7), on analyse k tableaux juxtaposés croisant les vaches ayant effectivement présenté k lactations (en lignes) et 8 variables décrivant la pathologie individuelle à l'échelle d'une lactation. On réalise ainsi successivement 7 analyses de tableaux juxtaposés et on obtient in fine 28 trajectoires sanitaires, définies par la longueur de la carrière et, pour une longueur de carrière déterminée, par les profils sanitaires individuels.

II.2.2. Les analyses sur variables instrumentales

Les analyses sur variables instrumentales sont des méthodes qui permettent d'expliquer un tableau de données par un autre tableau, intégrant plusieurs variables (appelées instrumentales) que l'on veut prendre en compte ou éliminer, globalement ou pour partie (**Chessel, 1995a**). Elles élargissent le concept de MANOVA aux tableaux de données jouant un rôle dissymétrique. Pour résumer le principe de ces analyses (ACPVI, AFCVI), on peut concevoir les analyses sur variables instrumentales comme l'ACP ou l'AFC de $\tilde{X}_{ik,j}$, projeté de $X_{ik,j}$ sur Y (**Obadia, 1978 - Sabatier et al, 1989, Lebreton et al., 1988**). Ces analyses utilisent donc des projecteurs, c'est à dire implicitement la métrique de Mahalanobis. L'intérêt de ces analyses et l'avantage qu'elles présentent par rapport aux analyses inter-classes et intra-classes qui n'intègrent qu'un seul effet, réside dans le fait qu'en « manipulant » les sous espaces engendrés par les indicatrices de classes liées aux différents effets (variables de Y), les données peuvent être explorées de façon organisée sous plusieurs points de vue. Avec deux variables instrumentales A et B , on distingue des effets simples A et B , des effets combinés $A + B$ (additif) et conjoint (AxB), des effets partiels (A sachant $B = A$ dans $A + B$, et B sachant A), et un effet interaction ($AxB - A + B$). (**Ter Braak, 1988 - Sabatier et al, 1989 - Chessel, 1995a**). Une même analyse décompose l'inertie comme le faisaient plusieurs analyses séparées (**Doledec et al, 1987**). Avec un seul effet, l'ACPVI ou l'AFCVI se ramènent à une analyse intra-classes (**Sabatier et al, 1989**). Nous illustrons cette méthode par deux exemples d'application, en épidémiologie, et en agronomie.

En épidémiologie, **Lescourret et al. (1991)** analyse un tableau croisant 181 lactations et 5 variables décrivant la qualité bactériologique du lait (notes d'abondance de 5 germes pathogènes), mis en relation avec un tableau de 5 variables potentiellement explicatives (qualitatives et quantitatives), dont une matérialisant l'année de prélèvement. Seules les 2 variables « exploitation » et « malpropreté » contribuent au premier plan factoriel, et sont retenues dans une deuxième AFCVI pour juger de l'effet exploitation (objectif de l'étude), conditionnellement à l'effet de malpropreté. Le tableau initial est mis en relation avec le tableau des résidus de la régression de la variable « exploitation » sur la variable « malpropreté », pondéré par les poids des lignes du tableau initial (effet partiel A sachant B).

Une ACPVI est utilisée par **Escoufier et al. (1988)** pour décrire un tableau croisant en lignes 78 vins rouges, en colonnes, 14 variables descriptives des composés phénoliques, associé à un tableau de variables indicatrices définissant le plan expérimental et constituées par le cépage (3 groupes), le millésime (1974 à 1985: 11 groupes) et la méthode de vinification (3 groupes). L'analyse est présentée comme un enchainement de décompositions de l'inertie par les classes du millésime en premier lieu, puis par les classes du cépage, sachant que l'effet du millésime a été enlevé, et enfin par l'interaction de ces deux effets. Un effet significatif du cépage dans la part d'inertie non expliquée par le millésime est mis en évidence.

II.2.3. L'analyse discriminante

Parce qu'elle génère des combinaisons linéaires de variables de variance unité, l'analyse discriminante classique constitue simultanément une analyse groupée qui maximise la variance inter-groupes et une analyse conditionnelle, qui minimise la variance intra-groupe. Elle est basée également sur l'utilisation de projecteur et s'apparente de fait aux analyses sur variables instrumentales. Cependant, un grand nombre de variables (par rapport au nombre de lignes) rend son emploi délicat et peu fréquent, comme c'est souvent le cas avec les variables taxonomiques (**Doledec et al, 1989**).

Les besoins d'analyse de tableaux évolutifs, notamment de la croissance, a fait émerger différentes méthodes que l'on peut rattacher à l'analyse discriminante, parmi lesquelles la méthode LONGI. Cette méthode a été développée par **Pernin (1986)** et appliquée à l'étude de données auxologiques chez l'enfant en réponse au problème induit par l'existence de données manquantes. L'auteur étudie 8 paramètres de croissance (J) chez 30 garçons et 30 filles (I), de la naissance à l'âge de 17 ans (K). On se place dans l'espace des observations x_{ijk} et non plus dans celui des individus, on transforme les variables « individus » et « âge » en indicatrices de classes qui définissent des sous espaces vectoriels $[I]$ et $[K]$. A chaque variable X_J on associe deux nouvelles variables centrée réduites, $X'_J = (X_J - \bar{X}_J)/s(X_J)$ avec \bar{X}_J moyenne de X_J sur l'ensemble des observations et $X''_J = (X_J - \bar{X}_{JK})/s(X_{JK})$ avec \bar{X}_{JK} moyenne de X_J pour l'âge K . On projette X'_J dans $[K] \cap [I]^\perp$ pour avoir un indice d'évolution, c'est-à-dire une combinaison linéaire des variables J qui simultanément maximise la corrélation avec K et minimise la corrélation avec I . On projette X''_J dans $[K]^\perp \cap [I]$ pour avoir un indice de situation qui maximise la corrélation avec I et

minimise la corrélation avec K . Dans l'exemple, 2 tranches d'âge particulières sont ainsi mises en évidence, 8-10 ans et 12-15 ans, et 4 trajectoires individuelles.

Le fait de définir ces analyses dans l'espace des observations autorise l'admission de données manquantes. Par rapport aux analyses sur variables instrumentales (projection sur $(A + B) \cap A^\perp$), l'originalité de la méthode et de ses variantes repose sur le fait qu'elle effectue simultanément une ACP inter-classes associée à un effet, et une ACP intra-classes associée à l'autre effet (**Yoccoz et al, 1988**), qui semble aboutir à leur meilleure séparation. La version AFC est décrite par **Pontier et al. (1987)**. Une autre variante, l'analyse discriminante de tableaux évolutifs, est donnée par **Casin (1995)**. D'autres méthodes ont été employées dans l'étude de Pernin (STATIS), que nous comparerons à la méthode LONGI dans le chapitre suivant.

Les analyses structurées que nous avons abordées au cours du chapitre II montrent à l'évidence de nombreux points communs, le principal d'entre eux étant d'analyser des tableaux projetés. Le choix de l'une d'entre elles dépendra du plan expérimental et des objectifs fixés (prise en compte ou l'élimination d'effets). Les analyses conjointes de plusieurs tableaux que nous envisageons maintenant répondent à un objectif différent: on ne cherche plus à décomposer l'inertie d'un tableau sur différentes variables, mais à décrire directement les structures induites dans les données par le plan expérimental (**Thioulouse et al., 1987**). Ces analyses ne substituent plus aux tableaux initiaux des projecteurs, mais des opérateurs d'inertie (**Chessel et al, 1996**).

III. Les analyses conjointes de K tableaux

Nous décrirons 3 méthodes d'analyses de K tableaux ayant été utilisées dans le traitement de tableaux indicés dans le temps: la méthode STATIS, l'analyse triadique partielle, et les analyses 3 modes.

III.1. La méthode STATIS

Cette méthode analyse une succession de K tableaux ayant au moins une marge en commun (I) et/ou (J), et comprend 3 étapes successives (**Lavit, 1988**). L'interstructure (première étape) recherche les ressemblances et dissemblances entre les K tableaux. A chaque tableau X_k (D matrice des poids des lignes, D_k matrice de poids des colonnes), on associe le

tableau des produits scalaires entre individus $X_k D_k X_k' D = W_k D$ (matrice de covariance de dimension $n \times n$, pour des données de départ centrées). La métrique choisie dans l'espace des opérateurs ($W_k D$) est la trace de $(W_k D W_L D)$, produit scalaire de $W_k D$ et $W_L D$. Un coefficient de corrélation est défini entre 2 tableaux:

$$Rv(X_K, X_L) = Tr(W_K D W_L D) / (\sqrt{Tr(W_K D W_K D)} \sqrt{Tr(W_L D W_L D)}).$$

On construit une matrice C contenant les produits scalaires entre opérateurs de rang K (ou les coefficients Rv si le nombre de variables diffèrent entre tableaux) sur laquelle on réalise une ACP, dont les axes définissent les plans factoriels de l'interstructure, et sur lesquels on positionne K points (avec l'aide à l'interprétation des Rv). L'étape du compromis consiste ensuite à trouver un opérateur « compromis » WD des K opérateurs $W_k D$, qui résume

l'ensemble des tableaux. Il est obtenu par $WD = \sum_{k=1}^K \alpha_k W_k D$, avec α_k composantes du

premier vecteur propre de C . Le compromis est donc une moyenne pondérée des K opérateurs $W_k D$, caractéristiques des K sous tableaux (les opérateurs sont traités comme le sont les individus dans une ACP classique). L'ACP du compromis WD fournit une représentation factorielle des pxk variables, des n individus, ainsi que des axes factoriels des analyses des sous tableaux (**Pernin, 1986**), permettant l'interprétation biologique de ces axes. L'intrastructure (troisième étape), décrit les écarts entre chaque individu et l'individu moyen (trajectoire).

On redonne ici à titre illustratif les conclusions de l'interstructure dans l'exemple de **Pernin (1986)**. L'interstructure montre une modification très importante de la croissance à partir de 13 ans, dues aux transformations pubertaires. Les points 4 à 13 ans sont chronologiquement ordonnés. Les structures des individus s'individualisent aux âges 7 et 11 ans (figure 4).

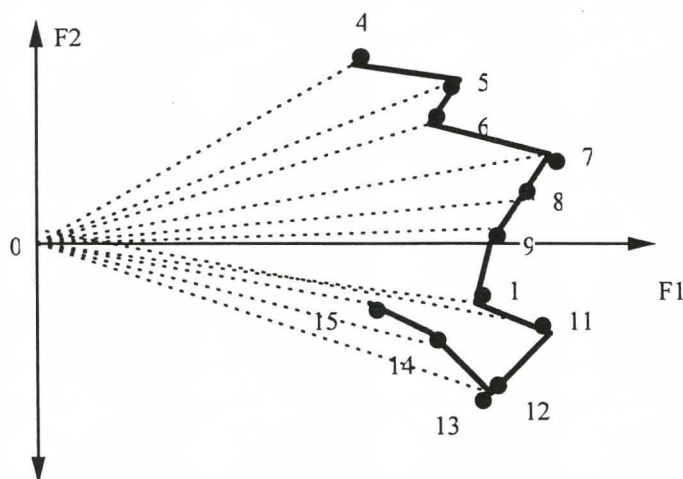


Figure 4.: Représentation du plan 1x2 de l'interstructure. Chaque point représente un âge (d'après **Pernin (1986)**)

La méthode STATIS présente 2 avantages, celui de conduire à une bonne synthèse des lignes des tableaux, et celui de permettre de réaliser des typologies de structure, et un défaut, la difficulté de représenter l'écart des tableaux séparés par rapport au compromis, au niveau des variables (**Chessel et al., 1995d**). Par rapport à STATIS, la méthode LONGI appliquée aux mêmes données donnent des résultats globalement identiques, tout en répondant à des objectifs différents. STATIS cherche à décrire l'évolution des relations entre tous les individus considérés pour toutes les variables mesurées, alors que LONGI cherche une expression multivariable de la croissance, indépendante des variations individuelles (**Pernin, 1986**). Outre l'objectif de l'étude, le choix entre l'une ou l'autre méthode dépendra de la nature des données: on utilisera LONGI en présence de données manquantes, et STATIS lorsque l'une des marges I ou J diffère pour les K tableaux.

Foucart (1978) fournit un exemple d'application de la méthode STATIS à l'étude d'une suite de tableaux de contingence. On associe à chacun des K tableaux de départ un tableau de Burt; et on applique STATIS à une suite de K tableaux de Burt (B_k), considérée comme une suite de matrice de variance-covariance (équivalent de $W_K D$). Le compromis B_C

s'écrit alors $B_C = \sum_{k=1}^K \alpha_k B_k$ avec $\sum_{k=1}^K \alpha_k = 1$.

III.2. L'analyse triadique partielle

Cette analyse est en fait une analyse STATIS sur tableaux par opposition à l'analyse STATIS sur opérateurs (matrice de variance-covariance) décrite précédemment. Elle permet de faire une analyse moyenne de plusieurs ACP (données quantitatives) portant sur les mêmes marges I et J (**Chessel et al., 1995c**). L'analyse comprend également 3 étapes. Les données des tableaux X_k sont centrées réduites et ré-organisées en un tableau $Z_{(I,J,K)}$, dont l'ACP non normée constitue l'interstructure. Pour chaque facteur obtenu, on ré-organise les coordonnées factorielle des lignes en une matrice « fictive » de dimension $n \times p$, et on fait l'ACP pour obtenir le compromis. L'intrastructure consiste à projeter en éléments supplémentaires sur le plan factoriel d'un compromis, les lignes du tableau $T_{(I,K,J)}$ pour obtenir la représentation des individus aux différentes dates et les lignes du tableau $T_{(I,J,K)}$ pour obtenir la représentation des variables aux différentes dates.

La méthode offre plusieurs avantages par rapport à STATIS. L'analyse triadique permet de considérer les données sous deux aspects différents, une suite de tableaux (individus-variables) indicés dans le temps ou une série de tableaux (dates-variables) indicés par individu. Dans l'interstructure, elle permet une description des individus par rapport à la typologie des dates. On obtient ensuite autant de compromis que de facteurs retenus à l'issue de l'interstructure, ce qui permet d'en affiner la description. Enfin, l'intrastructure permet d'obtenir aisément les trajectoires des individus et des variables, alors que STATIS ne permet de représenter que les trajectoires des individus. Enfin, sa simplicité d'exécution permet de transposer aux méthodes multi-tableaux les résultats classiques des analyses mono-tableau (analyse discriminante inter-tableaux, analyse canonique de plusieurs variables quantitatives) (**Thioulouse et al, 1987**). La méthode STATIS présente cependant l'avantage de pouvoir analyser K tableaux différant par l'une de leur marge.

En épidémiologie, **Lescourret (1994)** analyse ainsi une suite de 4 tableaux (correspondant à 4 mois successifs, juillet à octobre) croisant 40 élevages expérimentaux en lignes, 6 variables climatiques mensuelles en colonnes, l'objectif étant d'affecter à chaque élevage expérimental une « note » résumant son type de climat observé sur 4 mois, considéré dans la suite comme un facteur de risque potentiel des maladies d'élevage. L'axe 1 de l'interstructure exprime une ressemblance entre dates, et l'axe 2 une divergence entre août et octobre.

L'analyse de K tableaux de contingence pose le problème de la pondération des marges, différente d'un tableau à l'autre. **Foucart (1978)** propose 3 référentiels équivalents aux 3 étapes des deux analyses précédentes. Le premier référentiel (interstructure) s'obtient par l'AFC d'un tableau ($k \times n$) ayant en ligne les marges « lignes » des K tableaux. Le deuxième référentiel (compromis) est le résultat de l'AFC de $X = \frac{1}{K} \sum_{k=1}^K X_k$, moyenne des K tableaux, et le troisième référentiel (intrastructure) est obtenu en projetant en éléments supplémentaires les lignes et les colonnes des K tableaux initiaux sur le plan factoriel du compromis.

III.3. Les analyses 3-modes

Les analyses précédentes (STATIS, analyse triadique partielle) appartiennent à la famille des analyses « three-way two-mode » selon **Basford et al. (1991)**, dans le sens où une

des marges I ou J est privilégiée dans la construction du compromis. Ce choix peut être forcé par l'organisation des données (une des marges diffère entre les K tableaux) ou par les objectifs de l'étude (**Pernin, 1986**). Les analyses 3-modes appartiennent à la famille des analyses « three-way three-mode » pour lesquelles l'analyse de chacune des marges est effectuée simultanément (analyse symétrique d'un cube de données). Nous décrirons un modèle en particulier, le modèle de Tucker, fondement de l'ACP 3-mode, dont la méthode et les algorithmes sont détaillés par **Kroonenberg (1983)**.

Le modèle de Tucker généralisent l'ACP classique à des tableaux de dimensions trois, dans lesquels les marges I et J sont identiques pour les K répétitions. On réalise simultanément une ACP sur chacune des 3 marges. Par exemple, pour I , on diagonalise la matrice de covariance P de terme général $p_{ii'} = \sum_{j=1}^p \sum_{k=1}^K x_{ijk} x_{i'jk}$. On obtient in fine 3 matrices de vecteurs propres G , H , et E , soit 3 compromis, contenant respectivement p , q , et r vecteurs propres g , h , et e , retenus à l'issue de l'analyse, et reliés entre eux par $c_{pqr} = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K g_{ip} h_{jq} e_{kr} x_{ijk}$, terme général d'une matrice « noyau » (core matrix) de dimension $(pxqxr)$, équivalent 3D à la matrice diagonale de dimension 2 issue d'une ACP classique, contenant les racines carrées des valeurs propres $(\sqrt{\lambda_\alpha})$, et représentant le poids d'une combinaison linéaire de vecteurs propres de G , H , et E dans la reconstitution de x_{ijk} . En sommant les termes de cette matrice mis au carré, on obtient l'inertie totale expliquée par le modèle, la part d'inertie expliquée par chacun des 3 modes, et leur différentes interactions. L'originalité du modèle consiste dans le rôle symétrique joué par les 3 modes. Cependant, les données traitées étant rarement symétrique, plusieurs centrages des données initiales sont proposés, en fonction des objectifs, pour éliminer plusieurs effets additifs: des centrages par tranche par mode (ex j -centrage) et par bloc (jk -centrage). L'ACP 3-mode optimise les algorithmes de calcul par une procédure des moindres carrés alternés, basé sur un processus itératif répété jusqu'à convergence (**Kroonenberg, 1983**). D'autres modèles ont été proposés. Nous citerons parmi les principaux le modèle PARAFAC/CANDECOMP qui établit une matrice de vecteurs propres commune aux 3 modes (et non plus propre à chaque mode), et le modèle INDSCAL, dérivant du précédent lorsque les 2 premiers modes sont identiques (**Kroonenberg, 1983 - Lafaye, 1985 - Beffy, 1993 - Harshman et al., 1994**).

Les comparaisons des différentes méthodes analysant K tableaux sont peu nombreuses, en particulier entre les approches françaises (STATIS, analyse triadique) et les approches anglo-saxonnes (ACP 3 modes), et se limitent souvent aux aspects purement statistiques, comme la nature du critère optimisé (**Kiers, 1988**), sans référence aucune aux objectifs de l'analyse. Un exemple d'application de l'ACP 3-mode reprenant les données de **Thioulouse et al. (1987)** est donné par **Kroonenberg (1989)**. L'analyse envisage simultanément les données selon 3 points de vue: une série chronologique de tableaux [stations x variables], une série de tableaux [mois x variables] indicés par station, et une série de tableaux [mois x stations] indicés par variables. Elle conduit globalement au même compromis que l'analyse triadique, ce qui confirme la parenté entre cette dernière et le modèle de Tucker (**Kroonenberg, 1989, citant Jaffrenou, 1978**). D'un point de vue général, les analyses 3 modes reposent sur la notion d'automodélisation (ajustement direct des données), ou modélisation sans contraintes externes par opposition aux analyses structurées (**Beffy, 1993**). Ces analyses sont fondées sur la recherche de vecteurs propres communs ou spécifiques à chaque mode, alors que STATIS substituent aux tableaux initiaux des opérateurs d'inertie, ce qui lui donne un rôle relativement plus statistique et confirmatoire (**Kroonenberg, 1983**).

Toutes ces méthodes aboutissent à une description de la complexité des structures présentes dans un cube de données. Cependant, pas plus que les méthodes vues au chapitre II, ces analyses ne tiennent compte implicitement du caractère ordonné des données temporelles. Nous choisissons de regrouper dans cette quatrième partie, un ensemble de méthodes, assez disparates, mais aboutissant toute à une modélisation des séries temporelles.

IV. Les analyses de données temporelles

Sous ce titre général, figure la prise en compte de la notion de contiguïté temporelle, notion particulièrement intéressante lorsque les mesures ou les observations effectuées sur les mêmes individus I sont autocorrélées dans le temps, comme c'est le cas pour tout phénomène évolutif. Dans la situation présente, l'objectif consiste donc clairement, en restant au niveau de généralité 2 de **Pernin (1986)** à réaliser une typologie de chroniques temporelles. Ici encore, plusieurs méthodes peuvent être décrites, selon l'organisation des données et les objectifs poursuivis. La plus simple consiste à réaliser une ACP directement sur les données brutes.

IV.1. L'ACP sur données brutes

L'ACP peut être appliquée directement sur les données brutes si, d'un individu à l'autre, les conditions d'acquisition de données sont standardisées (tous les individus sont mesurés le même nombre de fois, aux mêmes dates), chaque date déterminant alors une variable (**Houllier, 1987**). Les exemples d'application sont nombreux, notamment lorsqu'il s'agit de décrire la croissance par un tableau multi-individus, multi-dates (**Moulin, 1993 - Granger et al., 1980 - Pernin, 1986 - Pardoux, 1989**). Cependant, dans ces analyses, **Houllier (1987)** pose la question du choix de la métrique, de sa justification face à des données qui ne sont que la discrétisation d'un phénomène continu sous jacent, ainsi que de la prise en compte de la variabilité intra-individuelle (représentée par l'autocorrélation des mesures successives), que l'ACP classique ne permet pas de prendre en compte. Pour intégrer l'ordre chronologique des mesures, 2 groupes de méthodes seront envisagées, les ACP sur graphe de voisinage et les ACP sur paramètres d'un modèle.

IV.2. L'analyse sur graphe de voisinage

Les méthodes que nous allons envisager maintenant généralisent encore davantage le point de vue des analyses structurées vues au chapitre II (**Escofier et al, 1988a**). L'information complémentaire apportée en marge ligne du tableau initial est une matrice de voisinage A (graphe de voisinage), qui pondère les lignes en tenant compte de la contiguité qui existe entre elles. Au tableau X_{ik} est associée une matrice $A_{(n,n)}$ de terme général $a_{ii'}$, qui vaut 1 si i et i' sont voisins ($i' \in V_{(i)}$), 0 sinon, i ne pouvant être voisin de lui même. Les lignes ainsi pondérées, la variance totale du tableau peut se décomposer d'une manière tout à fait similaire à celle vue dans les analyses structurées, en une variance locale et une variance globale (**Chessell et al., 1995b**):

$$Var = \sum_{i=1}^n p_i (x_i - \bar{x})^2 = \frac{1}{2} \sum_{i,j} p_i p_j (x_i - x_j)^2 = \frac{1}{2} \sum_{i \in V_{(i)}} p_i p_{i'} (x_i - x_{i'})^2 + \frac{1}{2} \sum_{i' \notin V_{(i)}} p_i p_{i'} (x_i - x_{i'})^2 .$$

Deux exemples en agronomie sont décrits. Pour modéliser les courbes de production fruitière de 20 mandariniers, **Méot et al. (1993)** couplent à un tableau X_{ij} croisant en ligne 15 années successives et en colonnes les 20 arbres, un graphe de voisinage définissant la structure de contiguité des années entre elles. Leur objectif est de trouver un modèle de la tendance (autocorrélations positives entre années) et de l'alternance (autocorrélations négatives entre années) dans l'évolution de la production des arbres. La diagonalisation d'un opérateur de

variance globale permet dans un premier temps d'isoler 2 vecteurs propres, sur lesquels on réalise la régression multiple des courbes initiales, donnant \tilde{X}_{IJ} . L'ACP de \tilde{X}_{IJ} suivi de la reconstitution des données à partir de son premier axe donne un premier modèle « lissé » du tableau de départ. On projette ensuite le tableau des résidus de la précédente analyse R_{IJ} sur les 5 premiers vecteurs propres issus de la diagonalisation de l'opérateur de variance locale, donnant \tilde{R}_{IJ} . L'ACP de \tilde{R}_{IJ} , suivi de la reconstitution des données sur les 2 premiers facteurs, conduit, après addition au modèle lissé, à un modèle prenant en compte simultanément, ce qui en fait son principal intérêt, la variance globale (analyse de tendance) et la variance locale (autocorrélation négative). Les auteurs soulignent l'équivalence entre ces méthodes et l'ACPVI du tableau initial couplé aux vecteurs propres de l'opérateur de voisinage (variables instrumentales).

Pour étudier l'évolution de la production agricole entre 1959 et 1985, et en particulier caractériser dans le temps des ensembles de départements ayant les mêmes profils agricoles, **Escofier et al. (1988a)** adoptent une approche voisine, sur un tableau $X_{(IK)J}$ croisant en ligne 89 départements (I) et 27 années (K), et 13 productions (J). 2 lignes sont voisines si, pour un même département, elles correspondent à 2 années contigües. Ils réalisent une analyse factorielle lissée en remplaçant chaque valeur initiale $x_{ik,j}$ par le barycentre de ses voisins $y_{ik,j}$ défini par $y_{ik,j} = \sum_{ik' \in V_{(ik)}} (x_{ik',j} / n_{ik})$, suivie d'une analyse factorielle des différences locales en remplaçant $x_{ik,j}$ par $x_{ik,j} - y_{ik,j}$. Quand la structure de contiguïté est définie par une partition de IK , la matrice de voisinage est composée de blocs de 1 sur sa diagonale, et l'analyse est ramenée à une analyse interclasses et intraclasse, comme celles décrites au chapitre II (**Le Foll, 1982 - Carlier, 1985**).

L'observation longitudinale de longue durée engendre souvent des données incomplètes (données manquantes) ou asynchrones (entre individus). Il est alors nécessaire comme le souligne **Abidi et al. (1995)** de standardiser les données après leur récolte, ce à quoi aboutit l'utilisation d'un modèle d'ajustement en préalable à l'analyse factorielle.

IV.3. L'ACP sur les paramètres d'un modèle.

Le modèle employé peut être un modèle paramétrique, en général non linéaire, utilisé avec une métrique particulière pour définir une distance entre courbes, ou un modèle linéaire,

comme les fonctions polynomiales. L'avantage commun à ces méthodes est de s'affranchir de la contrainte d'un plan expérimental hétérogène.

C'est probablement dans les études de croissance que ces modèles ont été le plus utilisés. **Pernin (1986)** comme **Abidi et al. (1995)** utilise pour modéliser les courbes de croissance d'enfants un modèle logistique, dépendant de 5 paramètres, auxquels ils appliquent une ACP en utilisant une métrique euclidienne particulière, la métrique de sensibilité, décrite par **Houllier (1987)**, qui tient compte de la forme de la courbe, et de l'impact de chacun des paramètres. L'avantage qu'ont ces méthodes sur les modèles polynomiaux et l'ACP directe, réside dans le fait que leur paramètres sont peu nombreux et souvent biologiquement interprétables, ce que constate **Abidi et al. (1995)** en comparant les cartes factorielles de l'ACP sur données brutes et l'ACP sur paramètres. L'inconvénient réside dans le fait que les courbes de croissance sont ajustées à partir d'un modèle commun, fixé a priori, avec une précision sur les paramètres variant d'un individu à l'autre, justifiant l'utilisation d'une métrique adéquate.

Champely (1994) utilise des fonctions B-splines pour ajuster par segment des séries temporelles et effectue une ACP sur les coefficients B-splines obtenus. Les splines sont des fonctions polynomiales par morceaux, qui permettent d'interpoler ou de lisser des données de forme quelconque contrairement au modèle déterministe préalablement envisagé. A partir des données brutes, on définit une base de fonctions B-spline $B_j^m(t)$, m étant le degré des fonctions polynomiales, et $j(1 \rightarrow m+k)$ avec k , le nombre de noeuds. La fonction spline qui ajuste les données s'obtient par combinaison linéaire des B-splines par $f(t) = \sum_{j=1}^{m+k} a_j B_j^m(t)$ où

a_j sont les coefficients B-splines. En gardant les n dates d'observations, on obtient une interpolation des données caractérisée par n coefficients (la courbe ajustée passe exactement par les points). Si au contraire les mesures contiennent du bruit, on ne garde que r dates ($r < n$) et on obtient une approximation (un lissage) de la courbe des données, caractérisée par r coefficients. L'ACP sur les coefficients a_j utilise, pour calculer une distance entre courbes, la métrique de Gram, produit scalaire généralisé aux fonctions continues, qui, selon qu'il intègre ou pas les dérivés, permet de faire apparaître les tendances longues ou les différences locales. L'ACP sur coefficients spline réalise donc une double automodélisation des séries temporelles. Par rapport aux méthodes classiques, elle a l'avantage de s'appliquer dans des situations limites: discrétisation irrégulière des données (dates et/ou espacements différents),

présence de données manquantes, processus non stationnaires. Elle permet également, en utilisant un centrage et une réduction des coefficients B-spline, de comparer des courbes non homogènes (évolution temporelle de variables différentes). Enfin, la modélisation de processus temporels par les B-splines peut se généraliser aux analyses de tableaux couplés (ACP sur variables instrumentales et analyse de co-inertie). D'autres méthodes d'interpolation et de lissage ont été utilisées, comme par exemple les fonctions I-splines (**El Faouzi et al., 1991 - Champely, 1994**) pour modéliser des courbes de croissance.

Conclusion

Que fait un biométricien avec un tableau de données individus x caractères x époques? **Dagnelie (1980)** répondait à cette question en privilégiant fortement les méthodes classiques d'analyse de variance multivariées et de regression linéaire et curvilinéaire. Comme le souligne **Champely (1994)**, cette approche cherche moins à décrire qu'à tester divers traitements. Or, les objectifs que nous nous fixons dans une approche écopathologique des troubles sanitaires nécessitent clairement une approche exploratoire, visant à décrire les structures d'un ensemble de données complexe (multivariées). De plus, les données sont très généralement acquises sur le terrain. On ne maîtrise pas alors l'ensemble des éléments du plan expérimental. Des animaux sont pesés à des dates différentes, prélevés à des âges différents, dans un contexte où les variations saisonnières, annuelles, ou physiologiques sont des facteurs explicatifs prépondérants. Aussi, l'approche inférentielle classique semble dans bien des cas mal adaptée à la prise en compte de la composante temporelle, ou non pertinente, dans le cas par exemple de séries non stationnaires, de moyenne et de variance non constante ... (**Champely, 1994**).

Au contraire, par la diversité des « options » qu'elle présente, l'analyse exploratoire montre une grande souplesse à s'adapter à diverses contraintes présentes dans les données (observations manquantes et asynchrones) et à différents objectifs fixés (élimination d'effet, identification d'une structure commune aux différentes dates, analyse de l'évolution d'une structure à différentes dates ...), que nous avons évoqués dans cette revue. On notera ici qu'en fine, les 2 approches classique et exploratoire ne sont pas en concurrence, mais bel et bien complémentaires, comme le montrent les analyses sur variables instrumentales.

Au total, un large éventail de méthodes se présente à l'utilisateur, parmi lesquelles un choix n'est pas toujours aisé. Chacune des méthodes pourra être prise en défaut, face à un

certain type de données, pour un certain objectif. La synergie méthode - données est un élément clé de l'analyse des données (Chessel et al., 1995e). Se pose alors le problème de la transposition d'une méthode d'une discipline à une autre. En effet, les différentes méthodes ont été développées indépendamment, par des équipes différentes, dans des disciplines différentes, auxquelles elles sont plus ou moins « adaptées ». Les comparaisons de méthodes sont peu nombreuses, et se limitent trop souvent à des considérations statistiques qui peuvent déconcerter l'utilisateur biologiste. Il apparaît donc essentiel d'objectiver le choix d'une méthode parmi un ensemble, sur la base d'un même jeu de données, comme dans l'exemple décrit par **Chessel et al. (1995f)**, dans lequel des méthodes comme STATIS et l'analyse de co-inertie multiple (**Chessel et al., 1996**), sont comparées sur la base d'un jeu de tableaux indicés dans le temps, alors qu'elles ne sont pas utilisées classiquement pour ce type de données.

Remerciements

Je remercie mon Directeur de Recherche et mes parrains scientifiques, pour leur aide et leurs conseils pertinents. Je remercie également le Directeur du CIRAD-EMVT et le Directeur du Laboratoire d'Ecopathologie de l'INRA, qui me permettent d'effectuer ce D.E.A. dans les meilleures conditions.

Bibliographie

- Abidi H., Pontier J., Borms J., Duquet W.**, 1995, Courbes de croissance: intérêt de la modélisation pour l'analyse de données longitudinales, *Rev. Stat. Appl.*, 43, 3, 55-72.
- Basford K.E., Kroonenberg P.M., DeLacy I.H.**, 1991, Three-way methods for multiattribute genotype x environment data: an illustrated partial survey, *Fields Crops Research*, 27, 131-157.
- Beffy J.L.**, 1993, Approche des structures ternaires en écologie par l'analyse en composantes principales à trois modes, thèse de Doctorat Universitaire, Université Claude Bernard Lyon I, 360 p.
- Carlier A.**, 1985, Application de l'analyse factorielle des évolutions et de l'analyse intra-période, *Stat. Ana. Donn.*, 10, 1, 27-53.
- Carlier A., Lavit C., Pages M., Pernin M.O., Turlot J. C.**, 1989, a comparative review of methods which handle a set of indexed data tables, In Coppi R., and Bolasco S. (Eds), *Multi-way data analysis*, 85-101.
- Casin Ph.**, 1995, L'analyse discriminante de tableaux évolutifs, *Rev. Stat. Appl.*, 43, 3, 73-91.

- Cazes P.**, 1982, Notes sur les éléments supplémentaires en analyse des correspondances. I. Pratique et utilisation [El. Supp.I], Les Cahiers de l'Analyse des Données, 1, 9-23.
- Cazes P.**, 1982, Notes sur les éléments supplémentaires en analyse des correspondances. II. Tableaux multiples [El. Supp.II], Les Cahiers de l'Analyse des Données, 2, 133-154.
- Cazes P., Chessel D., Doledéc S.**, 1988, L'analyse des correspondances internes d'un tableau partitionné: son usage en hydrobiologie, Rev. Stat. Appl., 36, 1, 39-54.
- Centofenti M., Chessel D., Doledéc S.**, 1989, Stabilité d'une structure spatiale et compromis d'une analyse statistique multi-tableaux: applications à la physico-chimie d'un lac réservoir, Rev. Fr. Sci. Eau, 2, 71-93.
- Champely S.**, 1994, Analyses multivariées de données fonctionnelles: approximation par les splines de regression, thèse de Doctorat, Université Claude Bernard Lyon I, 151 p.
- Chessel D.**, 1995a, Ordination sous contraintes, In Programmathèque ADE - 4, Analyses multivariées et expression graphique des données environnementales, fascicule 5, Variables instrumentales, Institut d'Analyse des Systèmes Biologiques et Socio-économiques, Université Claude Bernard Lyon I, 40 p.
- Chessel D., Thioulouse J.**, 1995b, Grille de mesures, In Programmathèque ADE - 4, Analyses multivariées et expression graphique des données environnementales, fascicule 3, , Ordination spatiales, Institut d'Analyse des Systèmes Biologiques et Socio-économiques, Université Claude Bernard Lyon I, 27 p.
- Chessel D., Thioulouse J., Champely S.**, 1995c, L'analyse triadique partielle, In Programmathèque ADE - 4, Analyses multivariées et expression graphique des données environnementales, fascicule 6, , Méthodes classiques d'ordination simultanée de k tableaux, Institut d'Analyse des Systèmes Biologiques et Socio-économiques, Université Claude Bernard Lyon I, 28 p.
- Chessel D., Thioulouse J., Hanafi M.**, 1995d, STATIS et valeurs typologiques des groupes faunistiques, In Programmathèque ADE - 4, Analyses multivariées et expression graphique des données environnementales, fascicule 6, , Méthodes classiques d'ordination simultanée de k tableaux, Institut d'Analyse des Systèmes Biologiques et Socio-économiques, Université Claude Bernard Lyon I, 29 p.
- Chessel D.**, 1995e, L'analyse factorielle multiple, In Programmathèque ADE - 4, Analyses multivariées et expression graphique des données environnementales, fascicule 6, , Méthodes classiques d'ordination simultanée de k tableaux, Institut d'Analyse des Systèmes Biologiques et Socio-économiques, Université Claude Bernard Lyon I, 26 p.
- Chessel D.**, 1995f, Instabilité de descripteurs et stabilité de structures en végétation aquatique, In Programmathèque ADE - 4, Analyses multivariées et expression graphique des

- données environnementales, fascicule 7, Co-inertie de k tableaux, Institut d'Analyse des Systèmes Biologiques et Socio-économiques, Université Claude Bernard Lyon I, 36 p.
- Chessel D., Hanafi M.**, 1996, Analyses de la co-inertie de K nuages de points, *Rev. Stat. Appl.*; à paraître.
- Dagnelie P.**, 1980, Une réponse à la question « Que fait un biométricien avec un tableau de données individus x caractères x époques »?, In Jolivet E. (Ed), *Biométrie et temps*, 52-57..
- Dakki M.**, 1986, Biotypologie et gradient thermique spatio-temporel, étude sur un cours d'eau du Moyen Atlas (Maroc), *Bull. Ecol.*, 17, 2, 79-85.
- Dessouter D.**, 1995, Les pneumopathies des petits ruminants au Sénégal: étude séro-épidémiologique et bio-géographique, thèse de Doctorat, Université Paris XII Val de Marne - Créteil, Volume I et II (annexes), 229 p.
- Doledec S., Chessel D.**, 1987, Rythmes saisonniers et composantes stationnelles en milieu aquatique. I. Description d'un plan d'observation complet par projection de variables, *Acta Oecologica, Oecol. Gener.*, 8, 3, 403-426.
- Doledec S., Chessel D.**, 1989, Rythmes saisonniers et composantes stationnelles en milieu aquatique: II. Prise en compte et élimination d'effets dans un tableau faunistique, *Acta Oecologica, Oecol. Gener.*, 10, 3, 207-232.
- Drouet D.**, 1983, Comparaison de tableaux de fréquence. Application à l'étude des évolutions des causes de décès en Bretagne, thèse de 3^{ème} cycle, Université Paris VI.
- El Faouzi N., Escoufier Y.**, 1991, Modélisation I-spline et comparaison de courbes de croissance, *Rev. Stat. Appl.*, 39, 1, 51-64.
- Escoufier B., Drouet D.**, 1983, Analyse des différences entre plusieurs tableaux de fréquence, *Les Cahiers de l'Analyse des Données*, 8, 4, 491-499.
- Escoufier B., Benali H.**, 1988a, Analyse factorielle lissée et analyse factorielle des différences locales, IRISA, Rennes, publication interne n°444, 34 p.
- Escoufier B., Pages J.**, 1988b, Analyses factorielles simples et multiples, objectifs, méthodes et interprétation, DUNOD, Paris, 241 p.
- Escoufier Y.**, 1980, L'analyse conjointe de plusieurs matrices de données, In Jolivet E. (Ed), *Biométrie et temps*, 59-76.
- Escoufier Y.**, 1985, Objectifs et procédures de l'analyse conjointe de plusieurs tableaux de données, *Stat. Ana. Donn.*, 10, 1, 1-10.

- Escoufier Y., Holmes S., 1988,** Decomposition de la variabilité dans les analyses exploratoires: un exemple d'analyse en composantes principales en présence de variables qualitatives concomitantes,
- Facy F., Govaert G., Laurent F., 1984,** Analyse typologique de données chronologiques en santé publique, *Rev. Stat. Appl.*, 32, 3, 43-63.
- Faye B., Brochart M., 1986a,** Enquête éco-pathologique continue: 7. Approche épidémiologique des mammites post-partum chez la vache laitière: étude descriptive et typologie des élevages, *Ann. Rech. Vét.*, 17, 3, 297-311.
- Faye B., Fayet J.C., Genest M., Chassagne M., 1986b,** Enquête écopathologique continue: 10. variations des fréquences pathologiques en élevage bovin laitier en fonction de la saison, de l'année et du numéro de lactation, *Ann. Rech. Vét.*, 17, 233-246.
- Faye B., Grelet Y., 1991,** Profils sanitaires en élevage bovin laitier: mise en relation avec une typologie d'exploitation, *Etudes et recherches sur les systèmes agraires et le développement*, n°21, 13-47.
- Faye B., Duc C., Landais E., Coulon J.B., Lescourret F., 1994a,** Types de trajectoires du statut sanitaire au cours de la carrière des vaches laitières, *Vet. Res.*, 25, 300-304.
- Faye B., Lefevre P.C., Lancelot R., Quirin R., 1994b,** Ecopathologie animale, méthodologie, applications en milieu tropical, CIRAD-EMVT / INRA, 115 p.
- Faye B., Boulme R., Lescourret F., Coulon J.B.,** A method for describing disease patterns during the life-span in dairy cows, *Vet. Res.*, à paraître.
- Foucart T., 1978,** Sur les suites de tableaux de contingence indexés par le temps, *Stat. Ana. Donn.*, 2, 67-85.
- Franquet E., Chessel D., 1994,** Approche statistique des composantes spatiales et temporelle de la relation faune-milieu, *C. R. Acad. Sci. Paris, Sciences de la vie*, 317, 202-206.
- Furse M.T., Moss D., Wright J.F., Armitage D., 1984,** The influence of seasonal and taxonomic factors on the ordination and classification of running - water sites in Great Britain and on the prediction of their macro-invertebrate communities, *Fresh. Biol.*, 14, 257-280.
- Gillet J., Madec F., Tillon J.P., 1987,** Etude longitudinale des états de santé et de performance observés dans des élevages de porcs en Bretagne, *Journées de la recherche porcine*, 19, 143-160.
- Gopalan T., 1980,** L'évolution du commerce d'exportation de l'Inde entre 1963 et 1975 [Export.Inde], *Les Cahiers de l'Analyse des Données*, 4, 407-442.
- Granger Y., Pages J., 1980,** Un exemple: les courbes de croissance d'organes du blé, In Jolivet E. (Ed), *Biométrie et temps*, 77-87.

- Harshman R.A., Lundy M.E.**, 1994, PARAFAC: parallel factor analysis, *Computational Statistics and Data Analysis*, 18, 39-72.
- Houllier F.**, 1987, Comparaison de courbes et modèles de croissance. Choix d'une distance entre individus, *Stat. Ana. Donn.*, 12, 3, 17-36.
- Kiers H.A.L.**, 1988, Comparison of « Anglo-saxon » and « French » three-mode methods, *Stat. Ana. Donn.*, 1988, 13, 3, 14-32.
- Kosmelj K., Lachet B.**, 1983, Aspect temporel des relations entre les variables hydriques du Haut Rhône français, *Rev. Stat. Appl.*, 31, 2, 5-18.
- Kroonenberg P.M.**, 1983, Three-modes principal component analysis, DSWO Press, Leiden, Pays Bas, 398 p.
- Kroonenberg P.M.**, 1989, The analysis of multiple tables in factorial ecology. III. - Three-mode principal component analysis: « analyse triadique complète », *Acta Oecologica, Oecol. Gener.*, 10, 3, 245-256.
- Lafaye J.Y.**, 1985, Application du modèle INDSCAL, *Stat. Ana. Donn.*, 10, 1, 83-102.
- Lavit Ch.**, 1988, Analyse conjointe de tableaux quantitatifs, Collection méthodes et programmes, Masson (Paris), 252 p.
- Lebreton J.D., Chessel D., Richardot-Coulet M., Yoccoz N.**, 1988, L'analyse des relations espèces-milieu par l'analyse canonique des correspondances. II. Variables de milieu quantitative, *Acta Oecologica, Oecol. Gener.*, 9, 2, 137-151.
- Le Foll Y.**, 1982, Pondération des distances en analyse factorielle, *Stat. Ana. Donn.*, 1, 13-31.
- Lescourret F. Faye B.**, 1991, Stratégie statistique du laboratoire d'écopathologie, *Epidémiol. et Santé Anim.*, 20, 103-105.
- Lescourret F.**, 1994, Modélisation des séries temporelles et rôle du graphisme, *Vet. Res.*, 25, 1140-146.
- Lopez C., Chatelin Y.M.**, 1991, Mammite, modélisation des taux cellulaires, étude de l'incidence des facteurs de conduite d'élevage sur l'évolution des taux, *Acta N1 Institut*, 81 p.
- Méot A., Chessel D., Sabatier R.**, 1993, Opérateurs de voisinage et analyse des données spatio-temporelles, In Lebreton J.D., Asselain B. (Eds), Masson, Paris, 45-71.
- Moulin C.H.**, 1993, Performances animales et pratiques d'élevage en Afrique Sahélienne: la diversité du fonctionnement des troupeaux de petits ruminants dans la communauté rurale de Ndiagne (Sénégal), thèse d'ingénieur, Institut National Paris - grignon, 243 p. et annexes.
- Mourad G.**, 1983, Flux de pétrole et flux de marchandises entre l'O.P.E.P. et l'O.C.D.E. de 1970 à 1979 [O.P.E.P. - O.C.D.E.], *Les Cahiers de l'Analyse des Données*, 1, 27-72.

- Nicholson M.J., Sayers A.R.**, 1987, Repeatability, reproducibility and sequential use of condition scoring of *Bos Indicus* cattle, *Trop. Anim. Hlth. Prod.*, 19, 127-135.
- Obadia J.**, 1978, L'analyse en composante explicative, *Rev. Stat. Appl.*, 26, 4, 5-28.
- Pardoux C.**, 1989, Apport de l'analyse factorielle à l'étude d'un processus, *Rev. Stat. Appl.*, 37, 4, 41-60.
- Pernin, M.O.**, 1986, Contribution à la méthodologie d'analyse de données longitudinales: exemple de la croissance chez l'être humain (auxologie), thèse de Doctorat, Université Claude Bernard Lyon I, 264 p.
- Persat H., Chessel D.**, 1989, typologie de distributions en classes de taille: intérêt dans l'étude des populations de poissons et d'invertébrés, *Acta Oecologica, Oecol. Gener.*, 10, 2, 175-195.
- Pontier J., Jolicoeur P., Pernin M.O.**, 1987, Analyse canonique complète, *Stat. Ana. Donn.*, 12, 1-2, 124-148.
- Sabatier R., Lebreton J.D., Chessel D.**, 1989, Principal component analysis with instrumental variables as a tool for modelling composition data, In Coopri R., Bolasco S. (Eds), *Multiway data analysis*, Elsevier Science, North Holland, 341-352.
- Swaine M.D., Greig-Smith P.**, 1980, An application of principal components analysis to vegetation change in permanent plots, *J. Ecol.*, 68, 33-41.
- Ter Braak C.J.F.**, 1988, Partial canonical correspondence analysis, In Bock H.H. (Ed), *Classification and related method of data analysis*, North Holland, 551-558.
- Thioulouse J., Chessel D.**, 1987, Les analyses multi-tableaux en écologie factorielle. I. - De la typologie d'état à la typologie de fonctionnement par l'analyse triadique, *Acta Oecologica, Oecol. Gener.*, 8, 4, 463-480.
- Tourrand J.F.**, 1994, L'élevage dans la révolution agricole au Waalo. Ruptures et continuité, thèse de Doctorat, Université Paris XII Val de Marne - Créteil, 416 p.
- Vincent B.**, 1983, Variations spatiales et saisonnières de la structures de groupements macrobenthiques littoraux en climat froid, *Hydrobiol.*, 102, 175-186.
- Vincent B., Vaillancourt G., Harvey M., Lafontaine N.**, 1983, Variations spatio-temporelles de la faune macro-benthique dans le Saint Laurent (Québec), *Arch. Hydriobiol.*, 98, 2, 181-197.
- Williams W.T., Stephenson W.**, 1973, The analysis of three-dimensional data (sites x species x times) in marine ecology, *J. Exp. Mar. Biol. Ecol.*, 11, 207-227.
- Yoccoz N., Chessel D.**, 1988, Ordination sous contraintes des relevés d'avifaune: éliminations d'effets dans un plan d'observations à deux facteurs, *C.R. Acad. Sci.*, 307, série III, 189-194.